Biometrika (2005), **92,** 3, *pp.* 691–701

Printed in Great Britain

Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach

By GUODONG LI and WAI KEUNG LI

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong

ligd@hkusua.hku.hk   hrntlwk@hku.hk

Summary

The recent paper by Peng & Yao (2003) gave an interesting extension of least absolute deviation estimation to generalised autoregressive conditional heteroscedasticity, GARCH, time series models. The asymptotic distributions of absolute residual autocorrelations and squared residual autocorrelations from the GARCH model estimated by the least absolute deviation method are derived in this paper. These results lead to two useful diagnostic tools which can be used to check whether or not a GARCH model fitted by using the least absolute deviation method is adequate. Some simulation experiments give further support to the asymptotic theory and a real data example is also reported.

Some key words: Absolute residual autocorrelation; Asymptotic distribution; Diagnostic checking; GARCH model; Local least absolute deviation estimator; Squared residual autocorrelation.

1. Introduction

With the practical motivation of modelling conditional heteroscedasticity in economic and financial time series, Engle (1982) and Bollerslev (1986) proposed respectively the autoregressive conditional heteroscedasticity, ARCH, and the generalised autoregressive conditional heteroscedasticity, GARCH, models. These two models have achieved huge success in real applications where an explicit conditional Gaussian likelihood function is often assumed to facilitate estimation of parameters in the model. Recently, however, more and more empirical evidence has suggested that financial time series can be very heavy-tailed (Mittnik et al., 1998; Rachev & Mittnik, 2000, Ch. 1). For this case traditional maximum likelihood estimation is not available, but the Gaussian likelihood function still can be used to estimate the parameters, corresponding to so-called quasi-maximum-likelihood estimation (Bollerslev & Wooldridge, 1992). Weiss (1986) obtained the asymptotic normality of quasi-maximum-likelihood estimators for the ARCH model. Hall & Yao (2003) established the asymptotic normality of the quasi-maximum-likelihood estimators for the general GARCH(p, q) models under certain conditions. They also discovered that the asymptotic distribution may not be normal, with an infinite fourth moment. Peng & Yao (2003) discussed three different local least absolute deviation estimators for the general GARCH(p, q) model and obtained their asymptotic distribution, which has a Gaussian form. Their result can be applied to the situation where the time series has infinite fourth moment, since conditions are imposed on the probability density function of the errors but not on the moments. Model estimation is only one of the stages in the Box–Jenkins approach to time series modelling. The last stage, checking whether or not a fitted GARCH model based on least absolute deviation estimation is adequate, has not been discussed in the literature.

Note that residuals coming from a correct time series model should be very close to white noise. Based on the asymptotic distribution of the residual autocorrelations, if the model is correct, we can derive tests for individual residual autocorrelations and also an overall portmanteau statistic for model diagnostic checking. For the asymptotic distribution of the residual autocorrelations for general nonlinear time series models, see Li (1992). This paper is motivated by Li & Mak (1994), who used the squared residual autocorrelations to devise some useful diagnostic tools for time series models with changing conditional variance. However the existence of squared residual autocorrelations needs a finite fourth moment, which excludes many heavy-tailed distributions. In this paper the asymptotic distribution of the autocorrelations of absolute residuals from a GARCH model fitted using the least absolute deviation method is derived. The result depends only on the existence of a second-order moment and is therefore robust under heavy-tailed distributions. This result allows us to construct a portmanteau statistic that is useful in checking model adequacy. In the finance literature absolute return autocorrelations, which are similar to absolute residual autocorrelations, have been discussed for example by Taylor (1986, pp. 52–6) and Granger et al. (1999). As a by-product a portmanteau test based on squared residual autocorrelations from the least absolute deviation fit is also derived in this paper.

2. Notation and definitions

Let $\{X_t\}$ be a stationary and ergodic time series generated by the following GARCH (p, q) model:

$$X_t = \eta_t \sigma_t, \quad \sigma_t^2 = c + \sum_{i=1}^{p} a_i X_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2, \tag{2·1}$$

where $c > 0$, $a_i \geqslant 0$ and $b_j \geqslant 0$ are unknown parameters, $\{\eta_t\}$ is a sequence of independent and identically distributed random variables with mean 0 and variance 1, and η_t is independent of $\{X_{t-k}, k \geqslant 1\}$ for all t. The necessary and sufficient condition for model (2·1) to exist as a unique strictly stationary process $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$ with $EX_t^2 < \infty$ is that

$$\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j < 1. \tag{2·2}$$

It is easy to find a positive value δ such that the median of ε_t^2 is equal to 1, where $\varepsilon_t = \eta_t/\delta$. Then we can rewrite model (2·1) in the form

$$X_t = \varepsilon_t h_t^{1/2}, \quad h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{2·3}$$

where $h_t^{1/2} = \delta \sigma_t$, $\alpha_0 = \delta^2 c$, $\alpha_i = \delta^2 a_i$ and $\beta_j = b_j$. Let $\theta = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^{\mathrm{T}}$, which is the parameter vector for model (2·3).

For any fixed t, $h_t(\theta)$ and $\varepsilon_t(\theta)=X_t/h_t^{1/2}(\theta)$ are functions of θ. Furthermore we define $Z_t(\theta)=\log \varepsilon_t^2(\theta)$ and $\operatorname{sgn}\{Z_t(\theta)\}$, where $\operatorname{sgn}(x)$ denotes the sign of x. Let

$$U_t(\theta)=\frac{\partial}{\partial\theta}\log h_t(\theta)=\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta},$$

which is also a function of θ. Let $\Sigma=E_0\{U_t(\theta^0)U_t(\theta^0)^{\mathrm{T}}\}$, a $(1+p+q)\times(1+p+q)$ matrix, where θ^0 is the true parameter value and E_0 denotes expectation under $\theta=\theta^0$.

For model (2·3), under (2·2) and other regularity conditions, Peng & Yao (2003) suggested a local least absolute deviation estimator $\hat{\theta}_{\mathrm{LAD}}$ for a given time series $\{X_{n-k}, k\geqslant 0\}$, defined by

$$\hat{\theta}_{\mathrm{LAD}}=\arg\min_{\theta\in\mathcal{N}}\sum_{t=1}^{n}|\log(X_t^2)-\log\{h_t(\theta)\}|, \tag{2·4}$$

where $\mathcal{N}$ denotes a sufficiently small but fixed neighbourhood of the true parameter value. In practice, we can suppose that $X_i=0$, for $i\leqslant -N$, where N is a large positive number, and Peng & Yao (2003) proved that such an estimation is as good as the $\hat{\theta}_{\mathrm{LAD}}$ obtained from (2·4), when N is large enough.

Based on the proof of Theorem 1 in Peng & Yao (2003), it can be proved that

$$\sqrt{n}(\hat{\theta}_{\mathrm{LAD}}-\theta^0)=\frac{\Sigma^{-1}}{2f(0)}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}U_t(\theta^0)\operatorname{sgn}\{Z_t(\theta^0)\}+o_p(1), \tag{2·5}$$

where $f(.)$ is the density function of $\log\varepsilon_t^2$. Hence, the first item on the right-hand side of (2·5) can be used to approximate $\sqrt{n}(\hat{\theta}_{\mathrm{LAD}}-\theta^0)$ in the derivation that follows.

Let $\hat{h}_t$ be the corresponding value when the parameter vector θ in h_t is replaced by $\hat{\theta}_{\mathrm{LAD}}$, and $\hat{\varepsilon}_t=X_t/\hat{h}_t^{1/2}$ is the corresponding standardised residual. Similarly to Li & Mak (1994), the lag-k standardised absolute residual autocorrelation can be defined as

$$\tilde{\rho}_k=\frac{\sum_{t=k+1}^{n}(|X_t|/\hat{h}_t^{1/2}-\bar{\varepsilon}^*)(|X_{t-k}|/\hat{h}_{t-k}^{1/2}-\bar{\varepsilon}^*)}{\sum_{t=1}^{n}(|X_t|/\hat{h}_t^{1/2}-\bar{\varepsilon}^*)^2}\quad(k=1,2,\ldots), \tag{2·6}$$

where $\bar{\varepsilon}^*=n^{-1}\sum|X_t|/\hat{h}_t^{1/2}$ and n is the sample size. Generally it is more difficult to discuss the asymptotic behaviour of $\tilde{\rho}_k$. However, if the model is correct, it can be shown that $\bar{\varepsilon}^*$ will converge to $\mu^*=E|\varepsilon_t|$ in probability. Hence we consider $\hat{\rho}_k$ instead of $\tilde{\rho}_k$, where

$$\hat{\rho}_k=\frac{\sum_{t=k+1}^{n}(|X_t|/\hat{h}_t^{1/2}-\mu^*)(|X_{t-k}|/\hat{h}_{t-k}^{1/2}-\mu^*)}{\sum_{t=1}^{n}(|X_t|/\hat{h}_t^{1/2}-\mu^*)^2}\quad(k=1,2,\ldots), \tag{2·7}$$

and $\bar{\varepsilon}^*$ in (2·6) is replaced by μ^*.

For the lag-k standardised squared residual autocorrelation, we use the same definition as Li & Mak (1994), namely

$$\tilde{r}_k=\frac{\sum_{t=k+1}^{n}(X_t^2/\hat{h}_t-\bar{\varepsilon})(X_{t-k}^2/\hat{h}_{t-k}-\bar{\varepsilon})}{\sum_{t=1}^{n}(X_t^2/\hat{h}_t-\bar{\varepsilon})^2}\quad(k=1,2,\ldots), \tag{2·8}$$

where $\bar{\varepsilon}=n^{-1}\sum X_t^2/\hat{h}_t$. Note that $\bar{\varepsilon}$ converges to $\mu=E\varepsilon_t^2$ in probability, if the model is correct. Hence, for the same reason as that for the absolute residuals, we consider $\hat{r}_k$ instead, where

$$\hat{r}_k=\frac{\sum_{t=k+1}^{n}(X_t^2/\hat{h}_t-\mu)(X_{t-k}^2/\hat{h}_{t-k}-\mu)}{\sum_{t=1}^{n}(X_t^2/\hat{h}_t-\mu)^2}\quad(k=1,2,\ldots) \tag{2·9}$$

and $\bar{\varepsilon}$ in (2·8) is replaced by μ.

In the following two sections, we will consider the asymptotic distributions of $(\hat{\rho}_1, \ldots, \hat{\rho}_M)^{\mathrm{T}}$ and $(\hat{r}_1, \ldots, \hat{r}_M)^{\mathrm{T}}$, where M is a positive integer.

3. Asymptotic distribution of the absolute residual autocorrelations

In this section we derive the asymptotic distribution of the vector $(\hat{\rho}_1, \ldots, \hat{\rho}_M)^{\mathrm{T}}$, which leads to a useful diagnostic tool for checking the adequacy of a GARCH model fitted by the least absolute deviation approach. For simplicity, from now on, we denote the local least absolute deviation estimator $\hat{\theta}_{\mathrm{LAD}}$ by $\hat{\theta}$ and the true parameter value θ^0 by θ.

We note that, if the model is correct, the item $n^{-1} \sum (|X_t|/\hat{h}_t^{1/2} - \mu^*)^2$ in (2·7) converges to the constant $(\sigma^*)^2 = \mathrm{var}(|\varepsilon_t|)$ in probability. Hence, for $\hat{\rho}_k$, we need only consider the asymptotic distribution of

$$\hat{C}_k^* = \frac{1}{n} \sum_{t=k+1}^{n} \left(\frac{|X_t|}{\hat{h}_t^{1/2}} - \mu^* \right) \left(\frac{|X_{t-k}|}{\hat{h}_{t-k}^{1/2}} - \mu^* \right).$$

Let $\hat{C}^* = (\hat{C}_1^*, \hat{C}_2^*, \ldots, \hat{C}_M^*)^{\mathrm{T}}$ and $C^* = (C_1^*, C_2^*, \ldots, C_M^*)^{\mathrm{T}}$, where C_k^* is the corresponding value when $\hat{\theta}$ in h_t is replaced by θ. Similarly define $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_M)^{\mathrm{T}}$ and $\rho = (\rho_1, \ldots, \rho_M)^{\mathrm{T}}$. By Taylor's expansion, we have

$$\hat{C}^* \simeq C^* + \frac{\partial C^*}{\partial \theta} (\hat{\theta} - \theta), \tag{3·1}$$

where $\partial C^*/\partial \theta = (\partial C_1^*/\partial \theta, \partial C_2^*/\partial \theta, \ldots, \partial C_M^*/\partial \theta)^{\mathrm{T}}$ and, for $k = 1, \ldots, M$,

$$\begin{aligned} \frac{\partial C_k^*}{\partial \theta} &= \frac{\partial}{\partial \theta} \left\{ \frac{1}{n} \sum_t \left(\frac{|X_t|}{h_t^{1/2}} - \mu^* \right) \left(\frac{|X_{t-k}|}{h_{t-k}^{1/2}} - \mu^* \right) \right\} \\ &= -\frac{1}{2n} \sum_t \frac{|X_t|}{h_t^{3/2}} \frac{\partial h_t}{\partial \theta} \left(\frac{|X_{t-k}|}{h_{t-k}^{1/2}} - \mu^* \right) - \frac{1}{2n} \sum_t \left(\frac{|X_t|}{h_t^{1/2}} - \mu^* \right) \frac{|X_{t-k}|}{h_{t-k}^{3/2}} \frac{\partial h_{t-k}}{\partial \theta}. \end{aligned}$$

By the ergodic theorem the second item on the right-hand side converges to zero. Hence, for large n, $\partial C_k^*/\partial \theta \simeq -\sum (|X_t|/h_t^{3/2})(\partial h_t/\partial \theta)(|X_{t-k}|/h_{t-k}^{1/2} - \mu^*)/(2n)$. If we take the conditional expectation of each term under the summation sign and apply the ergodic theorem, $\partial C_k^*/\partial \theta$ can be consistently estimated by

$$-\tilde{Y}_k^* = -\frac{1}{2n} \sum_t \frac{\mu^*}{h_t} \frac{\partial h_t}{\partial \theta} \left(\frac{|X_{t-k}|}{h_{t-k}^{1/2}} - \mu^* \right).$$

Define the resulting $M \times (1 + p + q)$ matrix by $-\tilde{Y}^*$ when $\partial C_k^*/\partial \theta$ in $\partial C^*/\partial \theta$ are replaced by $-\tilde{Y}_k^*$ $(k = 1, \ldots, M)$. Let H be the probability limit of $\tilde{Y}^*$. Then $\hat{C}^*$ in (3·1) can be approximated by

$$\hat{C}^* \simeq C^* - H(\hat{\theta} - \theta). \tag{3·2}$$

For the vector C^* on the right-hand side of (3·2), applying Theorem 2.8.1 in Lehmann (1998) directly, we can obtain that

$$\surd n C^* \to N\{0, (\sigma^*)^4 I_M\}, \tag{3·3}$$

in distribution, where I_M is the $M \times M$ identity matrix.

For the asymptotic covariance between $\sqrt{n}(\hat{\theta}-\theta)$ and $\sqrt{n}C^*$, replacing $\sqrt{n}(\hat{\theta}-\theta)$ with (2·5) first, we can obtain that

$$\begin{aligned}
\operatorname{cov}\{\sqrt{n}(\hat{\theta}-\theta), \sqrt{n}C_k^*\} &\simeq E\left[\frac{\Sigma^{-1}}{2f(0)}\frac{1}{\sqrt{n}}\sum_t U_t(\theta)\operatorname{sgn}\{Z_t(\theta)\}\sqrt{n}C_k^*\right] \\
&= E\left[\frac{\Sigma^{-1}}{2f(0)}\frac{1}{\sqrt{n}}\sum_t U_t(\theta)\operatorname{sgn}\{Z_t(\theta)\}\right. \\
&\qquad \left.\times \sqrt{n}\frac{1}{n}\sum_s\left(\frac{|X_s|}{\sqrt{h_s}}-\mu^*\right)\left(\frac{|X_{s-k}|}{\sqrt{h_{s-k}}}-\mu^*\right)\right] \\
&= \frac{\Sigma^{-1}}{2nf(0)}\sum_t\sum_s E\left[U_t(\theta)\operatorname{sgn}\{Z_t(\theta)\}\left(\frac{|X_s|}{\sqrt{h_s}}-\mu^*\right)\left(\frac{|X_{s-k}|}{\sqrt{h_{s-k}}}-\mu^*\right)\right] \\
&= \frac{\Sigma^{-1}}{2nf(0)}\sum_t E\left[U_t(\theta)\operatorname{sgn}\{Z_t(\theta)\}\left(\frac{|X_t|}{\sqrt{h_t}}-\mu^*\right)\left(\frac{|X_{t-k}|}{\sqrt{h_{t-k}}}-\mu^*\right)\right] \\
&= \frac{\Sigma^{-1}}{2f(0)}E\left[\left(\frac{|X_t|}{\sqrt{h_t}}-\mu^*\right)\operatorname{sgn}\{Z_t(\theta)\}\right]E\left\{U_t(\theta)\left(\frac{|X_{t-k}|}{\sqrt{h_{t-k}}}-\mu^*\right)\right\} \\
&= \frac{\Sigma^{-1}}{2f(0)}d^*\frac{2}{\mu^*}E\left\{\frac{\mu^*}{2h_t}\frac{\partial h_t}{\partial\theta}\left(\frac{|X_{t-k}|}{\sqrt{h_{t-k}}}-\mu^*\right)\right\} \\
&= \frac{\Sigma^{-1}}{f(0)}\frac{d^*}{\mu^*}H_{(k)},
\end{aligned} \qquad (3{\cdot}4)$$

where $\mu^* = E|\varepsilon_t|$, $d^* = E\{|\varepsilon_t|(I_{\{|\varepsilon_t|>1\}} - I_{\{|\varepsilon_t|<1\}})\}$ and $H_{(k)}$ is the kth row of the matrix $H = (H_{(1)}, \ldots, H_{(M)})^{\mathrm{T}}$.

The asymptotic normality of $\sqrt{n}(\hat{\theta}-\theta)$ has been shown in Peng & Yao (2003). Hence, by applying the Mann–Wald device, the martingale central limit theorem, (2·5) and (3·2) to (3·4), we know that $\sqrt{n}\hat{C}^*$ is asymptotically normally distributed with mean 0 and covariance matrix $(\sigma^*)^4 V_1$, where

$$V_1 = I_M + \frac{\mu^* - 8d^*f(0)}{4\mu^* f^2(0)(\sigma^*)^4}H\Sigma^{-1}H^{\mathrm{T}},$$

$(\sigma^*)^2 = \operatorname{var}(|\varepsilon_t|)$. Furthermore, $\sqrt{n}\hat{\rho}$ is also asymptotically normally distributed with mean 0 and covariance matrix V_1.

From the above, we can obtain the correct asymptotic standard errors for the absolute residual autocorrelations. In particular, when η_t follows the standard normal distribution, the quantity $\{\mu^* - 8d^*f(0)\}/\{4\mu^* f^2(0)\}$ is equal to $-0{\cdot}09$. For the t distribution this quantity can be shown to be negative; for example, it is $-0{\cdot}39$ for the t_3 distribution. Hence we know that the asymptotic standard errors are generally less than $1/\sqrt{n}$, which is usually regarded as a crude standard error in diagnostic checking. Our result implies that the test, using simply $1{\cdot}96/\sqrt{n}$, could be too conservative. For the extreme situation in which the conditional variance σ_t^2 in (2·1) is constant, that is $H = 0$, the standard errors of $\hat{\rho}_k$, for $k = 1, \ldots, M$, are equal to $1/\sqrt{n}$. These results are typical and consistent with the classical result; see Box & Pierce (1970) and Li & Mak (1994).

In general, V_1 is not an idempotent matrix, and then $n\hat{\rho}^{\mathrm{T}}\hat{\rho}$ does not follow a χ^2 distribution asymptotically. However, the statistic

$$Q(M) = n\hat{\rho}^{\mathrm{T}} V_1^{-1}\hat{\rho}$$

will be asymptotically distributed as χ^2_M if the model is correct. This quantity should be useful as a portmanteau statistic for checking the adequacy of GARCH models that are estimated with the least absolute deviation approach. In practice, we can obtain the exact values of $f(0)$, μ^* and d^* if the distribution of η_t is known. Otherwise we can use the value of $n^{-1}\sum|\hat{\varepsilon}_t|I_{\{|\hat{\varepsilon}_t|>1\}} - n^{-1}\sum|\hat{\varepsilon}_t|I_{\{|\hat{\varepsilon}_t|<1\}}$ to replace d^* and $\bar{\varepsilon} = n^{-1}\sum|\hat{\varepsilon}_t|$ to replace μ^*. For $f(0)$, we can use some method of density estimation, such as the kernel method or smoothing splines, to obtain $\hat{f}(x)$ first and then use $\hat{f}(0)$ to replace $f(0)$, since $\{\log(\hat{\varepsilon}_t^2)\}$ is very close to the independent identically distributed sequence when the model is correct. The entries of H and Σ in the definition of matrix V_1 can be replaced by the corresponding sample averages as in Li & Mak (1994). The constant $(\sigma^*)^4$ can be replaced by $(\hat{C}_0^*)^2$. Tse & Zuo (1997) considered the optimal choice of M for portmanteau tests proposed in Li & Mak (1994).

4. Asymptotic distribution of the squared residual autocorrelations

If the model is correct, under the condition $E\eta_t^4 < +\infty$, the term $n^{-1}\sum(X_t^2/\hat{h}_t - \mu)^2$ in (2·9) converges to $\sigma^2 = \mathrm{var}(\varepsilon_t^2)$ in probability. It is therefore enough to consider just the asymptotic distribution of

$$\hat{C}_k = \frac{1}{n}\sum_{t=k+1}^{n}\left(\frac{X_t^2}{\hat{h}_t} - \mu\right)\left(\frac{X_{t-k}^2}{\hat{h}_{t-k}} - \mu\right).$$

Let $\hat{C} = (\hat{C}_1, \hat{C}_2, \ldots, \hat{C}_M)^{\mathrm{T}}$, $C = (C_1, C_2, \ldots, C_M)^{\mathrm{T}}$, $\hat{R} = (\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M)^{\mathrm{T}}$ and $R = (r_1, r_2, \ldots, r_M)^{\mathrm{T}}$, where C_k and r_k are the corresponding values when the parameter estimator $\hat{\theta}$ in h_t is replaced by θ. By Taylor's expansion, we have

$$\hat{C} \simeq C + \frac{\partial C}{\partial \theta}(\hat{\theta} - \theta), \tag{4·1}$$

where $\partial C/\partial\theta = (\partial C_1/\partial\theta, \partial C_2/\partial\theta, \ldots, \partial C_M/\partial\theta)^{\mathrm{T}}$ and, for $k = 1, \ldots, M$,

$$\frac{\partial C_k}{\partial\theta} = -\frac{1}{n}\sum_t \frac{X_t^2}{h_t^2}\frac{\partial h_t}{\partial\theta}\left(\frac{X_{t-k}^2}{h_{t-k}} - \mu\right) - \frac{1}{n}\sum_t\left(\frac{X_t^2}{h_t} - \mu\right)\frac{X_{t-k}^2}{h_{t-k}^2}\frac{\partial h_{t-k}}{\partial\theta}.$$

As in § 3, for large n, we replace $\partial C_k/\partial\theta$ by

$$-\tilde{Y}_k = -\frac{1}{n}\sum_t \frac{\mu}{h_t}\frac{\partial h_t}{\partial\theta}\left(\frac{X_{t-k}^2}{h_{t-k}} - \mu\right) \quad (k = 1, \ldots, M)$$

and replace $\partial C/\partial\theta$ by the matrix $-\tilde{Y}$, where $\tilde{Y} = (\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_M)^{\mathrm{T}}$. Let Y be the probability limit of $\tilde{Y}$. Then $\hat{C}$ in (4·1) can also be approximated by

$$\hat{C} \simeq C - Y(\hat{\theta} - \theta). \tag{4·2}$$

For the vector C on the right-hand side of (4·2), under the condition $E\eta_t^4 < +\infty$, we also have that

$$\sqrt{n}C \to N(0, \sigma^4 I) \tag{4·3}$$

in distribution.

For the asymptotic covariance between $\surd n(\hat{\theta}-\theta)$ and $\surd nC$, we have, by using (2·5), that

$$\begin{aligned}\operatorname{cov}\{\surd n(\hat{\theta}-\theta), \surd nC_k\} &\simeq E\left[\frac{\Sigma^{-1}}{2f(0)}\frac{1}{\surd n}\sum_t U_t(\theta)\operatorname{sgn}\{Z_t(\theta)\}\surd nC_k\right] \\ &= E\left[\frac{\Sigma^{-1}}{2f(0)}\frac{1}{\surd n}\sum_t U_t(\theta)\operatorname{sgn}\{Z_t(\theta)\}\right. \\ &\qquad \left.\times \surd n\frac{1}{n}\sum_s\left(\frac{X_s^2}{h_s}-\mu\right)\left(\frac{X_{s-k}^2}{h_{s-k}}-\mu\right)\right] \\ &= \frac{\Sigma^{-1}}{2f(0)}d\frac{1}{\mu}E\left\{\frac{\mu}{h_t}\frac{\partial h_t}{\partial\theta}\left(\frac{X_{t-k}^2}{h_{t-k}}-\mu\right)\right\} \\ &= \frac{\Sigma^{-1}}{2f(0)}\frac{d}{\mu}Y_{(k)}, \end{aligned} \qquad (4\cdot4)$$

where $\mu = E\varepsilon_t^2$, $d = E\{\varepsilon_t^2(I_{\{|\varepsilon_t|>1\}} - I_{\{|\varepsilon_t|<1\}})\}$ and $Y_{(k)}$ is the kth row of the matrix $Y = (Y_{(1)}, \ldots, Y_{(M)})^{\mathrm{T}}$.

By the Mann–Wald device, the martingale central limit theorem and $E\eta_t^4 < +\infty$, we know, from (4·2) to (4·4), that $\surd n\hat{C}$ is asymptotically normal with mean 0 and covariance matrix $\sigma^4 V_2$, and that $\surd n\hat{R}$ is also asymptotically normal with mean 0 and covariance matrix V_2, where

$$V_2 = I_M + \frac{\mu - 4df(0)}{4\mu f^2(0)\sigma^4} Y\Sigma^{-1}Y^{\mathrm{T}}.$$

Remark 1. As a result of the condition $E\eta_t^4 < +\infty$, we have no asymptotic result for the squared residuals when $\eta_t \sim t_3$ or $\eta_t \sim t_4$.

Remark 2. The quantity $\{\mu - 4df(0)\}/\{4\mu f^2(0)\}$ is positive when η_t follows the standard normal or t distribution; for example it is 1·44 for the normal distribution and 1·58 for the t_5 distribution. Then we know that the asymptotic standard errors are generally larger than $1/\surd n$ and our result implies that the test, using simply $1{\cdot}96/\surd n$, is too sensitive. This result is somewhat different from the absolute-value version in § 3 and the classical situation where the use of $1{\cdot}96/\surd n$ leads to under-rejection at the 5% significance level (Box & Pierce, 1970).

Remark 3. Again we use the corresponding sample averages to replace the entries of Y and Σ in the covariance matrix V_2. In general the constant σ^4 can be replaced by $\hat{C}_0^2$. Even under heavy-tailed distributions, we can still evaluate the covariance matrix V_2 this way. For other nuisance parameters, μ, d and $f(0)$, if the distribution of η_t is unknown, we can also use $\bar{\varepsilon} = n^{-1}\sum \hat{\varepsilon}_t^2$ to replace μ, $n^{-1}\sum \hat{\varepsilon}_t^2 I_{\{|\varepsilon_t|>1\}} - n^{-1}\sum \hat{\varepsilon}_t^2 I_{\{|\varepsilon_t|<1\}}$ to replace d and $\hat{f}(0)$ to replace $f(0)$, where $\hat{f}(.)$ is the density function estimated from the series $\{\log(\hat{\varepsilon}_t^2)\}$.

Similarly to § 3, V_1 is generally not an idempotent matrix, so that $n\hat{R}^{\mathrm{T}}\hat{R}$ does not follow a χ^2 distribution asymptotically. However, the portmanteau statistic

$$Q^2(M) = n\hat{R}^{\mathrm{T}}V_2^{-1}\hat{R}$$

will be asymptotically χ_M^2 if the model is correct. This quantity can also be used to check the adequacy of GARCH models fitted by the least absolute deviation method.

5. EMPIRICAL RESULTS AND DISCUSSION

For this section, we performed two simulation experiments to demonstrate the usefulness of the asymptotic results obtained in §§ 3 and 4. In the first experiment, we considered the GARCH(1, 1) time series

$$X_t = \eta_t \sigma_t, \quad \sigma_t^2 = 0{\cdot}4 + 0{\cdot}4 X_{t-1}^2 + 0{\cdot}1 \sigma_{t-1}^2,$$

where $\{\eta_t\}$ was an independent identically distributed sequence with mean 0, generated from the $N(0, 1)$, t_3 and t_5 distributions in three subexperiments. The η_t's were then scaled to have variance 1. For these three cases, we considered two different sample sizes, $n = 200$ and $n = 500$, and there were 500 independent replications for each model and sample size combination. The iteration algorithm of Nelder & Mead (1965), which is available in the International Mathematical and Statistical Library subroutine BCPOL, was used to search for the estimates which minimise (2·4). The subroutine BCPOL was also used in the second experiment and a real data example below. The asymptotic standard errors A_i $(i = 1, \ldots, 6)$ of the absolute residual autocorrelations, $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_6)^{\mathrm{T}}$, and the squared residual autocorrelations, $\hat{r} = (\hat{r}_1, \ldots, \hat{r}_6)^{\mathrm{T}}$, were computed according to the results in §§ 3 and 4. The empirical standard errors S_i of $\tilde{\rho}_i$ and $\tilde{r}_i$ $(i = 1, \ldots, 6)$ were also obtained and were taken to be the 'true' standard errors. Table 1 presents the empirical standard errors and the averages of the asymptotic standard errors for lags 1, 2, 3 and 6. It can be seen that the asymptotic results for both absolute and squared residual autocorrelations match the 'true' values satisfactorily for n as small as 200 and quite well for $n = 500$.

Table 1: *First simulation study. The empirical, S_i, and the large-sample, A_i, standard errors of absolute residual autocorrelations and squared residual autocorrelations for the* GARCH(1, 1) *model, for lags* 1, 2, 3 *and* 6

			lag i							
			Absolute residuals				Squared residuals			
Distribution	n		1	2	3	6	1	2	3	6
t_3	200	A_i	0·0679	0·0688	0·0701	0·0707				
		S_i	0·0640	0·0670	0·0665	0·0680				
	500	A_i	0·0432	0·0437	0·0444	0·0447				
		S_i	0·0433	0·0436	0·0414	0·0438				
t_5	200	A_i	0·0683	0·0692	0·0702	0·0706	0·0820	0·0803	0·0744	0·0715
		S_i	0·0712	0·0691	0·0724	0·0683	0·0752	0·0747	0·0701	0·0682
	500	A_i	0·0433	0·0438	0·0444	0·0447	0·0494	0·0488	0·0463	0·0449
		S_i	0·0431	0·0446	0·0454	0·0443	0·0419	0·0429	0·0412	0·0411
$N(0, 1)$	200	A_i	0·0693	0·0699	0·0704	0·0707	0·0911	0·0858	0·0767	0·0719
		S_i	0·0688	0·0669	0·0700	0·0702	0·0777	0·0721	0·0726	0·0718
	500	A_i	0·0439	0·0442	0·0446	0·0447	0·0561	0·0531	0·0477	0·0449
		S_i	0·0423	0·0433	0·0457	0·0422	0·0468	0·0443	0·0437	0·0422

In the second experiment, we considered the empirical size and power of the statistics $Q(M)$ and $Q^2(M)$. Two generating processes were involved: one is an ARCH process,

$$X_t = \eta_t \sigma_t, \quad \sigma_t^2 = c + a_1 X_{t-1}^2 + a_2 X_{t-2}^2 + a_3 X_{t-3}^2, \tag{5·1}$$

and the other is a GARCH process,

$$X_t = \eta_t \sigma_t, \quad \sigma_t^2 = c + a_1 X_{t-1}^2 + a_2 X_{t-2}^2 + b_1 \sigma_{t-1}^2, \tag{5·2}$$

where the standard normal distribution and the t_3 and t_5 distributions were again considered for the innovation sequence $\{\eta_t\}$. Again, the η_t's were scaled to have variance 1. We used two different sets of parameters, $\theta=(c, a_1, a_2, a_3)=(0{\cdot}4, 0{\cdot}2, 0{\cdot}4, 0)$ or $\theta=(0{\cdot}4, 0{\cdot}2, 0{\cdot}4, 0{\cdot}2)$ for the ARCH process and $\theta=(c, a_1, a_2, b_1)=(0{\cdot}4, 0{\cdot}4, 0, 0{\cdot}1)$ or $\theta=(0{\cdot}4, 0{\cdot}4, 0{\cdot}2, 0{\cdot}1)$ for the GARCH process. For each set of parameters, three different sample sizes, $n=200$, $n=500$ and $n=1000$, were considered. There were 1000 replications for each combination of θ, n and the distribution of η_t. In order to investigate sizes and powers of $Q(M)$ and $Q^2(M)$ with $M=6$, we estimated the simulated data under the assumption that $a_3=0$ for the ARCH process and $a_2=0$ for the GARCH process, and computed the values of $Q(M)$ and $Q^2(M)$. Although there is no theoretical result when $\eta_t \sim t_3$, we still can compute the value of $Q^2(M)$ empirically because of Remark 3. Table 2 displays proportions of rejections based on the upper 5th percentile of the corresponding asymptotic χ^2_6 distribution. Note that, when the generating process (5·1) or (5·2) is rewritten in the form (2·3), the parameter $\alpha_3=0{\cdot}2$ or $\alpha_2=0{\cdot}2$ will change to 0·091 for the normal distribution, 0·063 for the t_5 distribution and 0·039 for the t_3 distribution. Accordingly the powers in Table 2 decrease as the distribution of η_t becomes more heavy tailed. It can be seen that $Q^2(M)$ still has about the right size for the t_3 distribution when $n=200$. For larger sample sizes, $n=500$ and $n=1000$, the sizes of $Q(M)$ are all very close to 0·05 and are actually better than those of $Q^2(M)$ overall. All the powers of $Q(M)$ are larger than those of $Q^2(M)$. This dominance of $Q(M)$ over $Q^2(M)$ in terms of power is all the more remarkable under the heavy-tailed t distributions. This suggests that $Q(M)$ is the superior test.

Table 2: *Second simulation study. The empirical size and power of $Q(M)$ and $Q^2(M)$, based on 1000 replications and $M=6$*

		Size		Power		Size		Power	
Distribution	n	$Q(M)$	$Q^2(M)$	$Q(M)$	$Q^2(M)$	$Q(M)$	$Q^2(M)$	$Q(M)$	$Q^2(M)$
		ARCH model				GARCH model			
t_3	200	0·036	0·046	0·170	0·126	0·055	0·053	0·123	0·096
	500	0·051	0·060	0·390	0·200	0·052	0·044	0·258	0·107
	1000	0·051	0·058	0·631	0·242	0·049	0·056	0·391	0·141
t_5	200	0·039	0·047	0·224	0·178	0·046	0·043	0·173	0·111
	500	0·047	0·055	0·562	0·348	0·052	0·049	0·375	0·199
	1000	0·049	0·054	0·860	0·548	0·049	0·055	0·598	0·291
$N(0,1)$	200	0·041	0·046	0·281	0·276	0·043	0·027	0·201	0·142
	500	0·056	0·055	0·665	0·654	0·053	0·034	0·452	0·365
	1000	0·052	0·053	0·946	0·937	0·053	0·043	0·804	0·684

As a real application, we considered the simple rate of daily return, as a percentage, of the Hong Kong Hang Seng Index during (1993–2002). There were 2471 observations and an autoregressive model was considered first:

$$X_t = 0{\cdot}0376X_{t-1} - 0{\cdot}0417X_{t-2} + 0{\cdot}0954X_{t-3} - 0{\cdot}0508X_{t-4} - 0{\cdot}0369X_{t-5} + e_t.$$

Figure 1 shows the histogram of the residuals $\hat{e}_t$ corresponding to the above model. It suggests that the residuals $\hat{e}_t$ are more heavy tailed than the normal distribution. This fact is supported by the fact that the sample kurtosis of the $\hat{e}_t$ is approximately 8. Thus the maximum likelihood estimate or quasi-maximum-likelihood estimate is not suitable. The least absolute deviation method was used to estimate the following ARCH model

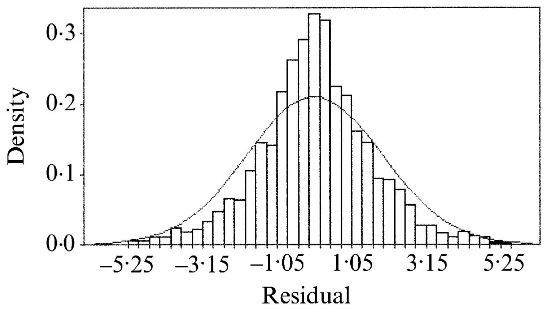

Fig. 1: Hang Seng Index data. The histogram of the residuals $\hat{e}_t$ and the normal density with the same mean and variance.

for the residuals $\hat{e}_t$:

$$e_t = \varepsilon_t h_t^{1/2}, \quad h_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j e_{t-j}^2,$$

where $E\varepsilon_t = 0$ and the median of ε_t^2 was equal to one. Note that here ε_t was not assumed to be normally distributed. We considered least absolute deviation estimation for three models with $p = 6$, $p = 7$ and $p = 8$. The initial value for α_0 was 1·6 and the initial value for the other parameters was 0·01. The methodology for diagnostic checking in §§ 2–4 was applied to these three models. We set M to be 10 and the values for μ, μ^*, d, d^*, σ^2, $(\sigma^*)^2$ and $f(0)$ were estimated with the methods mentioned in §§ 3 and 4. The bandwidth was set to be 0·1 for the estimation of $f(.)$. Our major interest concerns which of the three models can fit the data adequately. Table 3 presents the absolute residual autocorrelations $\tilde{\rho}_k$, the squared residual autocorrelations $\tilde{r}_k$ and their asymptotic standard errors A_k, where $k = 1, 2, 4, 8$. The complete table is available upon request. The overall test statistics $Q(10)$ and $Q^2(10)$ are also recorded in the same table. Based on the upper 5% significance level of the χ^2_{10} distribution, the ARCH(8) model fits the data adequately according to both statistics $Q(M)$ and $Q^2(M)$, and all individual absolute residual autocorrelations and

Table 3. *Model diagnostic checking results for the daily simple rate of return of the Hong Kong Hang Seng Index* (1993–2002). *Absolute residual autocorrelations, $\tilde{\rho}_k$, squared residual autocorrelations, $\tilde{r}_k$, and with standard errors in parentheses*

	$\rho = 6$		$p = 7$		$p = 8$	
k	$\tilde{\rho}_k$	$\tilde{r}_k$	$\tilde{\rho}_k$	$\tilde{r}_k$	$\tilde{\rho}_k$	$\tilde{r}_k$
1	0·0307	0·0330	0·0309	0·0346	0·0195	0·0254
	(0·0198)	(0·0221)	(0·0205)	(0·0228)	(0·0254)	(0·0271)
2	−0·0295	−0·0284	−0·0265	−0·0255	−0·0265	−0·0241
	(0·0198)	(0·0218)	(0·0206)	(0·0225)	(0·0255)	(0·0264)
4	−0·0263	−0·0267	−0·0208	−0·0233	−0·0022	−0·0100
	(0·0198)	(0·0216)	(0·0205)	(0·0223)	(0·0251)	(0·0261)
8	0·0675	0·0464	0·0645	0·0464	−0·0169	−0·0242
	(0·0201)	(0·0202)	(0·0202)	(0·0202)	(0·0248)	(0·0255)
$Q(10)$	33·5144		24·2346		8·1166	
$Q^2(10)$		20·9291		16·4843		7·3396

squared residual autocorrelations are nonsignificant, if the correct asymptotic critical value $1{\cdot}96A_i$ is used. For the ARCH(7), both residual autocorrelations at lag 8 are significantly different from zero using the critical value $1{\cdot}96A_i$. The model is correctly rejected by $Q(M)$. However, this model is regarded as adequate according to $Q^2(M)$, which is consistent with the observation that the portmanteau statistic $Q^2(M)$ is less powerful under a heavy-tailed situation. The ARCH(6) model is inadequate according to both criteria and there are four absolute residual autocorrelations and one squared residual autocorrelation that are significantly different from zero according to the critical value $1{\cdot}96A_i$.

Acknowledgement

W. K. Li thanks the Croucher Foundation for awarding a Senior Research Fellowship, a Hong Kong Research Grant Council grant for partial support and Dr S. Ling for comments and discussion. We are grateful to the editor and the referees for comments that led to improvement of the paper.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Economet.* **31**, 307–27.
Bollerslev, T. & Wooldridge, J. M. (1992). Quasi maximum likelihood estimation and inference in dynamic models with time varying covariance. *Economet. Rev.* **11**, 143–72.
Box, G. E. P. & Pierce, D. A. (1970). Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. *J. Am. Statist. Assoc.* **65**, 1509–26.
Engle, R. F. (1982). Autoregression conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987–1008.
Granger, C. W. J., Spear, S. & Ding, Z. X. (1999). Stylized facts on the temporal and distributional properties of absolute returns: an update. In *Statistics and Finance: an Interface*, Ed. W. S. Chan, W. K. Li and H. Tong, pp. 97–120. London: Imperial College Press.
Hall, P. & Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**, 285–317.
Lehmann, E. L. (1998). *Elements of Large-Sample Theory*. New York: Springer.
Li, W. K. (1992). On the asymptotic distribution of residual autocorrelations in nonlinear time series modelling. *Biometrika* **79**, 435–7.
Li, W. K. & Mak, T. K. (1994). On the squared residual autocorrelations in nonlinear time series with conditional heteroskedasticity. *J. Time Ser. Anal.* **15**, 627–36.
Mittnik, S., Rachev, S. T. & Paolella, M. S. (1998). Stable paretian models in finance: some empirical and theoretical aspects. In *A Practical Guide to Heavy Tails*, Ed. R. J. Adler, R. E. Feldman and M. S. Taqqu, pp. 79–110. Boston: Birkhauser.
Nelder, J. A. & Mead, R. (1965). A simplex method for function minimization. *Comp. J.* **7**, 308–13.
Peng, L. & Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967–75.
Rachev, S. T. & Mittnik, S. (2000). *Stable Paretian Models in Finance*. New York: Wiley.
Taylor, S. (1986). *Modelling Financial Time Series*. New York: John Wiley & Sons.
Tse, Y. K. & Zuo, X. L. (1997). Testing for conditional heteroscedasticity: some Monte Carlo results. *J. Statist. Comp. Simul.* **58**, 237–53.
Weiss, A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Economet. Theory* **2**, 107–31.

[*Received May* 2004. *Revised December* 2004]